SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 15157.7-MA	2. GOVT ACCESSION NO. N/A A117576	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) On the Local Convergence of Quasi-Newton Methods for Constrained Optimization		5. TYPE OF REPORT & PERIOD COVERED Reprint
		6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) Paul T. Boggs Jon W. Tolle Pyng Wang		8. CONTRACT OR GRANT NUMBER(s) DAAG29 77 G 0125
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of North Carolina Chapel Hill, NC 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE Mar 82
		13. NUMBER OF PAGES 11
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Submitted for announcement only.

DTIC ELECTE JUL 27 1982 S D B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A117576

DTIC FILE COPY

ARO 15157.7-MA

SIAM J. CONTROL AND OPTIMIZATION
Vol. 20, No. 2, March 1982

0363-0129/82/2002-0002 $01.00/0

ON THE LOCAL CONVERGENCE OF QUASI-NEWTON METHODS FOR CONSTRAINED OPTIMIZATION*

PAUL T. BOGGS†, JON W. TOLLE‡ AND PYNG WANG§

Abstract. We consider the application of a general class of quasi-Newton methods to the solution of the classical equality constrained nonlinear optimization problem. Specifically, we develop necessary and sufficient conditions for the Q-superlinear convergence of such methods and present a companion linear convergence theorem. The essential conditions relate to the manner in which the Hessian of the Lagrangian function is approximated.

1. Introduction. In this paper we consider means of solving the equality constrained nonlinear optimization problem

$$\text{(NLP)} \qquad \begin{aligned} &\text{Minimize } f(x) \\ &\text{subject to } g(x)=0, \end{aligned}$$

where it is assumed that $f:\mathbb{R}^n \to \mathbb{R}$ and $g:\mathbb{R}^n \to \mathbb{R}^m$ are smooth functions.

After two decades of experimentation with penalty function techniques, augmented Lagrangian functions, gradient projection methods and other procedures, research on numerical methods for solving NLP has recently centered on implementing some form of a quasi-Newton technique for this constrained problem. The preeminence of quasi-Newton methods for solving unconstrained nonlinear problems and good experimental results to date lead one to believe that this approach is sound. However, there remain numerous questions concerning convergence, rates of convergence, update formulas, and implementation that are as yet unanswered. It is the purpose of this paper to shed light on some of these questions, in particular, on the local and Q-superlinear convergence of these methods.

We define a quasi-Newton method for NLP as an iterative scheme which generates sequences $\{x^k\}$, $\{\lambda^k\}$, and $\{B_k\}$ from formulas

$$\lambda^{k+1} = \Lambda(x^k, \lambda^k, B_k), \tag{1.1}$$

$$B_k\delta_x^k = -l_x(x^k, \lambda^{k+1}), \tag{1.2}$$

$$x^{k+1} = x^k + \alpha^k\delta_x^k, \tag{1.3}$$

$$B_{k+1} = \mathcal{B}(x^k, x^{k+1}, \lambda^k, \lambda^{k+1}, B_k), \tag{1.4}$$

where x^0, λ^0 and B_0 are given, Λ and $\mathcal{B}$ are appropriate update functions and $l(x,\lambda) = f(x) + \lambda^T g(x)$ is the standard Lagrangian function. The step lengths α^k are obviously important, but for local convergence theory $\alpha^k = 1$ is the optimal choice and α^k will be taken to have this value throughout.

* Received by the editors February 6, 1980 and in final revised form March 27, 1981. This work was supported in part by the U.S. Army Research Office under grant DAAG29-79-G0014.

† U.S. Army Research Office, Research Triangle Park, North Carolina 27709 and Curriculum in Operations Research and Systems Analysis, University of North Carolina, Chapel Hill, North Carolina 27514.

‡ Department of Mathematics and Curriculum in Operations Research and Systems Analysis, University of North Carolina, Chapel Hill, North Carolina 27514.

§ Bell Laboratories, Whippany, New Jersey 07981.

Much of the recent work on quasi-Newton methods for NLP can be put into this framework. Powell [8], [9], following the work of Han [6], obtains δ_x^k by solving a quadratic program

$$\begin{aligned}&\text{Minimize } \nabla f(x^k)^T\delta_x^k + \tfrac{1}{2}\delta_x^{k\,T} B_x \delta_x^k\\ &\text{subject to } \nabla g(x^k)^T\delta_x^k = -g(x^k),\end{aligned}$$

and chooses λ^{k+1} to be the optimal multiplier vector for this program. B_k is a standard rank two update approximation to l_{xx} with a modification which assures that the B_k remain positive definite.

Tapia [10], [11] shows that Powell's choices of δ_x^k and λ^{k+1} can be obtained by applying a structured quasi-Newton method to the system

$$l_x(x,\lambda) = 0, \qquad l_\lambda(x,\lambda) = 0. \tag{1.5}$$

That is, x^{k+1} and λ^{k+1} are obtained from the equations

$$\begin{bmatrix} B_k & \nabla g(x^k) \\ \nabla g(x^k)^T & 0 \end{bmatrix}\begin{bmatrix} \delta_x^k \\ \delta_\lambda^k \end{bmatrix} = \begin{bmatrix} -l_x(x^k,\lambda^k) \\ -g(x^k) \end{bmatrix}, \tag{1.6}$$

$$\lambda^{k+1} = \lambda^k + \delta_\lambda^k, \tag{1.7}$$

with x^{k+1} given by (1.3). Here again B_k is an approximation to l_{xx}, so that the $(n+m)\times(n+m)$ matrix in (1.6) is a structured approximation to the Jacobian matrix for the system (1.5). It is easily seen that the solutions to (1.6), (1.7), given by

$$\lambda^{k+1} = (\nabla g(x^k)^T B_k^{-1}\nabla g(x^k))^{-1}\{g(x^k) - \nabla g(x^k)^T B_k^{-1}\nabla f(x^k)\}, \tag{1.8}$$

$$\begin{aligned}\delta_x^k = &-B_k^{-1}\{I - \nabla g(x^k)(\nabla g(x^k)^T B_k^{-1}\nabla g(x^k))^{-1}\nabla g(x^k)^T B_k^{-1}\}\nabla f(x^k)\\ &- B_k^{-1}\nabla g(x^k)(\nabla g(x^k)^T B_k^{-1}\nabla g(x^k))^{-1} g(x^k),\end{aligned} \tag{1.9}$$

also satisfy (1.1) and (1.2). In addition to the formula (1.8), Tapia presents a number of other possible updates for λ, preferring, for theoretical reasons, a double update of λ.

In [1] the authors have considered a variation of the system (1.5) in which the Lagrangian function $l(x,\lambda)$ is replaced by a more general Lagrangian $M(x,\lambda)$ which is quadratic in λ. The purpose for introducing this generalization was to obtain better convergence from poor starting points. Locally, however, the quasi-Newton equations derived from $M(x,\lambda)$ are nearly identical to those of (1.6).

The local convergence of these methods has been investigated by a number of authors. Before reviewing their results, we point out the distinction between Q-superlinear and R-superlinear rates of convergence and the difference between the convergence rates of the vector $\{(x^k,\lambda^k)\}$ and [illegible]onent $\{x^k\}$. Recall that a vector sequence $\{v^k\}$ converges R-superlinearly to v [illegible]nly if the sequence $\{\|v^k - v^*\|\}$ is bounded by a sequence which converges Q-su[illegible]inearly to zero. Because an R-superlinearly convergent sequence need not be even Q-linearly convergent, R-superlinear convergence by itself is computationally meaningless. It is also the case that the Q-superlinear convergence of $\{v^k\}$ implies only the R-superlinear convergence of its components. (See Tapia [10, section 8] for a more detailed discussion.) Since λ^{k+1} depends only on x^k and not on λ^k, to be most effective the structured quasi-Newton method should yield Q-superlinear convergence of the sequence $\{x^k\}$.

The major convergence analyses center on how well and in what sense the B_k generated by (1.4) approximate the Hessian of the Lagrangian function at (x^*,λ^*), the optimal solution pair. These analyses are based on similar studies of the unconstrained problem. In the latter case extensive use is made of the Broyden–Dennis–Moré analysis of the quasi-Newton update formulas and the Dennis–Moré characterization of Q-superlinear convergence. (See Dennis and Moré [4] for a survey of these results.) This characterization shows that Q-superlinear convergence in the unconstrained case occurs if and only if

$$\frac{|(\bar{B}_k - \nabla^2 f(x^*))\delta^k|}{|\delta^k|} \to 0, \tag{1.10}$$

where $\nabla^2 f(x^*)$ is the Hessian of the function to be minimized, δ^k is the step generated, and $\bar{B}_k$ is the approximate Hessian.

For the constrained case Powell [9] develops a procedure for updating the B_k and shows that, under the second order sufficiency conditions, the resulting method is at least R-superlinearly convergent in x. He also provides a condition related to (1.10) which is sufficient for "2-step" superlinear convergence. In particular, for the projection matrix

$$P(x) = I - \nabla g(x)(\nabla g(x)^T \nabla g(x))^{-1} \nabla g(x)^T,$$

Powell shows that

$$\frac{|P(x^k)(B_k - l_{xx}(x^*, \lambda^*))P(x^k)\delta_x^k|}{|\delta_x^k|} \to 0 \tag{1.11}$$

is sufficient for

$$\frac{|x^{k+1} - x^*|}{|x^{k-1} - x^*|} \to 0.$$

Powell was not able to show that his method satisfies this condition, however.

Also under the second order sufficiency conditions, Han [6] demonstrates the Q-superlinear convergence of $\{(x^k, \lambda^k)\}$ when a form of Greenstadt's update is used in (1.4). However, Han requires the stronger assumption that $l_{xx}(x^*, \lambda^*)$ be positive definite in order to obtain the Q-superlinear convergence in (x, λ) for the BFGS update. It should be noted that Greenstadt's method is not computationally attractive, since it almost always performs poorly in spite of its theoretical properties. To guarantee that $l_{xx}(x^*, \lambda^*)$ is positive definite requires the addition of a penalty term to the Lagrangian, a computationally unattractive option.

Tapia [10], [11] and Glad [5] obtain Q-superlinear convergence in (x, λ) for $l_{xx}(x^*, \lambda^*)$ positive definite. Tapia [10] obtains the stronger result of Q-superlinear convergence in x but at the cost of an additional update of λ at each step.

In this paper we first characterize Q-superlinear convergence in x for these methods (Theorem 3.1). The characterization is a natural generalization of the Dennis–Moré result (1.10). Simply put, it states that Q-superlinear convergence in x occurs if and only if

$$\frac{|P(x^k)(B_k - l_{xx}(x^*, \lambda^*))\delta_x^k|}{|\delta_x^k|} \to 0, \tag{1.12}$$

where $P(x)$ is the projection matrix given above. Note that (1.12) does not contain a post-multiplication of $B_k - l_{xx}(x^*, \lambda^*)$ by $P(x^k)$ as does (1.11), and hence, it takes into account the action of $B_k - l_{xx}(x^*, \lambda^*)$ off of the null space of $\nabla g(x^k)^T$, which (1.11) does not.

Using the characterization (1.11), we then show (Theorem 3.2) that Q-superlinear convergence in x is obtained when $l_{xx}(x^*, \lambda^*)$ is positive definite. This is a slightly stronger result than those previously published and reviewed above. Finally, a sufficient condition for Q-linear convergence in x is established (Theorem 3.3). This theorem also makes use of the matrices $P(x^k)(B_k - l_{xx}(x^*, \lambda^*))$, requiring that they be small in norm for all k. Hence it provides a complementary result to Theorem 3.1.

2. Basic notation and assumptions. For the problem NLP considered here we assume f and g are at least three times continuously differentiable and that the gradient $\nabla g(x)$ has full rank for all x. In addition we assume that NLP has a (local) solution x^* at which the second order sufficiency conditions hold. That is, there exists a unique vector $\lambda^* \in \mathbb{R}^m$ such that

(i) $l_x(x^*, \lambda^*) = 0$,
(ii) $\nabla g(x^*)^T y = 0$, $y \neq 0$ implies $y^T l_{xx}(x^*, \lambda^*) y > 0$.

For functions $h: \mathbb{R}^n \to \mathbb{R}^q$, we denote the Jacobian and Hessian matrices by $\nabla h(x)$ and $\nabla^2 h(x)$, respectively. Here, for notational convenience, $\nabla h(x)$ is always written as an $n \times q$ matrix. For functions of x and λ, we denote derivatives with respect to x or λ by subscripts; hence, $l_x(x, \lambda) = \nabla f(x) + \nabla g(x)\lambda$, $l_{x\lambda}(x, \lambda) = \nabla g(x)$, etc.

Vectors are always column vectors unless transposed, the transposition operation for vectors and matrices being indicated by a superscript T.

In the theory of constrained minimization, the projection of vectors onto the tangent space of the level sets of the constraints plays an important role. For a given $\hat{x}$, the matrix

$$P(\hat{x}) = [I - \nabla g(\hat{x})(\nabla g(\hat{x})^T \nabla g(\hat{x}))^{-1} \nabla g(\hat{x})^T]$$

projects vectors onto the tangent space of the smooth manifold

$$\{x : g(x) = g(\hat{x})\}$$

at $x = \hat{x}$. The projection onto the orthogonal complement of this tangent space will be denoted by $Q(\hat{x})$. Thus

$$Q(\hat{x}) = I - P(\hat{x}).$$

$|\cdot|$ will everywhere denote the l_2-norm. In § 3, it is necessary to use the Frobenius norm for matrices. The Frobenius norm weighted by the matrix M is denoted by $\|\cdot\|_M$.

3. Necessary and sufficient conditions for superlinear convergence. We consider the algorithm obtained by applying a structured quasi-Newton method to the system (1.5), thus obtaining the formulas given in (1.6) and (1.7) with solutions (1.8) and (1.9). This algorithm has the important property that δ_x^k satisfies the linearized constraints, i.e.,

$$g(x^k) + \nabla g(x^k)^T \delta_x^k = 0.$$

Extending the analysis by Powell [9], we obtain a necessary and sufficient condition for Q-superlinear convergence in x, given linear convergence and a few basic assumptions on the approximating matrices B_k. The essential condition is that the matrix B_k must approximate the Hessian matrix $l_{xx}(x^*, \lambda^*)$ in the sense of Dennis and Moré [3] but only when projected onto the tangent hyperplane to the surface $\{z : g(z) = g(x^k)\}$.

We assume in the remainder of this section that the B_k are symmetric, nonsingular, and uniformly bounded. In addition, we assume that the matrices B_k are uniformly positive definite on the null space of $\nabla g(x^*)^T$. That is, there exists a $\beta > 0$ such that whenever $y \neq 0$ and $\nabla g(x^*)^T y = 0$,

$$y^T B_k y \geqq \beta |y|^2$$

for every k. Thus we require the B_k to satisfy the second order sufficiency condition satisfied by $l_{xx}(x^*, \lambda^*)$ (see § 2). This assumption is slightly weaker than that of Powell who assumes that the B_k are positive definite on $\mathbb{R}^n$ and uniformly positive definite on the null space of $\nabla g(x^*)^T$.

Note that with the above assumptions on the matrices B_k, the matrices $\nabla g(x^*)^T B_k^{-1} \nabla g(x^*)$ are nonsingular. For if $\nabla g(x^*)^T B_k^{-1} \nabla g(x^*) z = 0$, then $B_k^{-1} \nabla g(x^*) z$ is in the null space of $\nabla g(x^*)^T$, and hence, $z \neq 0$ implies the contradiction $0 < (B_k^{-1}\nabla g(x^*)z)^T B_k (B_k^{-1} \nabla g(x^*)z) = z^T \nabla g(x^*)^T B_k^{-1} \nabla g(x^*) z = 0$. It follows that (1.8) and (1.9) are well-defined for x^k sufficiently near to x^*.

Our first result is a generalization of a result of Powell [9].

LEMMA 1. *The value of δ_x^k is invariant under the transformation*

$$B_k \to B_k + \nabla g(x^k) U^T \equiv \hat{B}_k,$$

where U is any $n \times m$ matrix such that both $\hat{B}_k$ and $\nabla g(x^k)^T \hat{B}_k^{-1} \nabla g(x^k)$ are nonsingular.

Proof. It follows from (1.9) that the lemma is true if the matrices

$$A_1 = \hat{B}_k^{-1} \nabla g(x^k)[\nabla g(x^k)^T \hat{B}_k^{-1} \nabla g(x^k)]^{-1}$$

and

$$A_2 = \hat{B}_k^{-1} - \hat{B}_k^{-1} \nabla g(x^k)[\nabla g(x^k)^T \hat{B}_k^{-1} \nabla g(x^k)]^{-1} \nabla g(x^k)^T \hat{B}_k^{-1}$$

are independent of U. The assumptions on U allow the use of the Sherman–Morrison–Woodbury formula (see, e.g., Ortega and Rheinboldt [7, p. 50]) to express $\hat{B}_k^{-1}$ as

$$\hat{B}_k^{-1} = B_k^{-1} - B_k^{-1} \nabla g(x^k)[I - U^T B_k^{-1} \nabla g(x^k)]^{-1} U^T B_k^{-1}.$$

Substitution of this expression into A_1 yields

$$A_1 = B_k^{-1} \nabla g(x^k)[\nabla g(x^k)^T B_k^{-1} \nabla g(x^k)]^{-1},$$

which establishes the result for A_1. For A_2, note that

$$A_2 = [I - A_1 \nabla g(x^k)] \hat{B}_k^{-1}.$$

Again using the expression for $\hat{B}_k^{-1}$ yields the desired result.

It follows from the assumptions made on the B_k that if $U = \gamma \nabla g(x^k)$, where γ is a sufficiently large positive constant, then the hypotheses of the lemma are satisfied. It should also be noted that the value of λ^{k+1} is not invariant under the given transformation in B_k. Thus, a variety of choices of λ^{k+1} give rise to the same value of δ_x^k (as demonstrated by Tapia in [11]). However, it is easily seen that the first order necessary conditions and equation (1.8) imply that if $\{x^k\} \to x^*$ then $\{\lambda^{k+1}\} \to \lambda^*$.

For convenience, we now write (1.9) in the form

$$-B_k \delta_x^k = V_k \nabla f(x^k) + W_k g(x^k). \tag{3.1}$$

We note that the two vectors on the right-hand side are conjugate with respect to B_k^{-1}; in fact, $V_k^T B_k^{-1} W_k = 0$. Letting P_k be the projection matrix at x^k defined in § 2, we see that

$$P_k V_k = P_k, \tag{3.2a}$$

$$V_k P_k = V_k, \tag{3.2b}$$

$$P_k W_k = 0. \tag{3.2c}$$

The next lemma is also a modification of the results of Powell. Two positive sequences, $\{s^k\}$ and $\{r^k\}$, which converge to zero are said to be of the same order if there exist positive constants c_1 and c_2 such that for k sufficiently large,

$$c_1 \leqq \frac{|r^k|}{|s^k|} \leqq c_2.$$

LEMMA 2. *Suppose $\{x^k\} \to \{x^*\}$ with a linear rate of convergence. Then the sequences $\{|\delta_x^k|\}$, $\{|x^k - x^*|\}$, and $\{|g(x^k)| + |P_k \nabla f(x^k)|\}$ converge to zero and are of the same order.*

Proof. Using Lemma 1 and the properties of the B_k, we see that by choosing γ sufficiently large we may replace the B_k by $\hat{B}_k$ for which $\hat{B}_k$ and $\hat{B}_k^{-1}$ are uniformly bounded and positive definite. The change does not affect the value of δ_x^k or the relations (3.2). Now using (3.2b) and (3.1) there exists an $\alpha_1 > 0$ such that

$$|\delta_x^k| \leqq \alpha_1\{|g(x^k)| + |P_k \nabla f(x^k)|\}.$$

(3.1), (3.2a), (3.2c), and the linearized constraint equation yield the existence of an $\alpha_2 > 0$ such that

$$\{|g(x^k)| + |P_k \nabla f(x^k)|\} \leqq \alpha_2 |\delta_x^k|.$$

Thus, $\{|\delta_x^k|\}$ and $\{|g(x^k)| + |P_k \nabla f(x^k)|\}$ are of equivalent order. That $\{|\delta_x^k|\}$ and $\{|x^k - x^*|\}$ are of the same order follows from the consequence of linear convergence

$$1 - r \leqq \frac{|\delta_x^k|}{|x^k - x^*|} \leqq 1 + r,$$

where $r < 1$. This completes the proof.

Now let $\{G_k\}$ be any sequence of matrices satisfying

(i) $$G_k \delta_x^k = l_x(x^{k+1}, \lambda^{k+1}) - l_x(x^k, \lambda^{k+1}),$$

(ii) $$G_k \to l_{xx}(x^*, \lambda^*).$$

For example, G_k could be chosen as

$$G_k = \int_0^1 l_{xx}(x^k + t\delta_x^k, \lambda^{k+1})\, dt.$$

LEMMA 3. *Assume $\{x^k\} \to x^*$ linearly. Then there exists an $\alpha > 0$ such that*

$$|x^{k+1} - x^*| \leqq \alpha[|\delta_x^k|^2 + |P_k(G_k - B_k)\delta_x^k|].$$

Proof. By Lemma 2 there exists an $\eta > 0$ such that

$$|x^{k+1} - x^*| \leqq \eta\{|g(x^{k+1})| + |P_{k+1} \nabla f(x^{k+1})|\}. \tag{3.3}$$

Now

$$g(x^{k+1}) = g(x^k) + \nabla g(x^k)^T \delta_x^k + O(|\delta_x^k|^2) = O(|\delta_x^k|^2). \tag{3.4}$$

From (i) above and (1.3),

$$(G_k - B_k)\delta_x^k = l_x(x^{k+1}, \lambda^{k+1}).$$

Using the fact that $P_{k+1} \nabla g(x^{k+1}) = 0$, we obtain the identity

$$(P_{k+1} - P_k)(G_k - B_k)\delta_x^k + P_k(G_k - B_k)\delta_x^k = P_{k+1} \nabla f(x^{k+1}). \tag{3.5}$$

The smoothness assumptions on $g(x)$ assure that

$$P_{k+1} - P_k = O(|\delta_x^k|) \tag{3.6}$$

and the lemma follows from (3.3)–(3.6) and the uniform boundedness of the G_k and the B_k.

We can now state and prove the necessary and sufficient conditions for Q-superlinear convergence for the structured quasi-Newton method as applied to the system (1.5).

THEOREM 3.1. *Let* $(\delta_x^k, \delta_\lambda^k)$ *satisfy* (1.6) *where the matrices* B_k *satisfy the conditions stated at the beginning of the section. Suppose* $\{x^k\} \to x^*$ *linearly. Then* $\{x^k\} \to x^*$ *Q-superlinearly if and only if*

$$\lim_{k\to\infty} \frac{|P_k(B_k - l_{xx}(x^*, \lambda^*))\delta_x^k|}{|\delta_x^k|} = 0, \tag{3.7}$$

where $P_k = I - \nabla g(x^k)(\nabla g(x^k)^T \nabla g(x^k))^{-1} \nabla g(x^k)^T$.

Proof. Let $\{G_k\}$ be a sequence of approximations to $l_{xx}(x^*, \lambda^*)$ as defined above. Clearly G_k can replace $l_{xx}(x^*, \lambda^*)$ in (3.7). Now suppose (3.7) holds. Then by Lemma 3

$$|x^{k+1} - x^*| = o(|\delta_x^k|).$$

But by Lemma 2 $\{|\delta_x^k|\}$ and $\{|x^k - x^*|\}$ are of the same order; hence there is a constant $\alpha > 0$ such that

$$\frac{|x^{k+1} - x^*|}{|x^k - x^*|} \leqq \alpha \cdot \frac{|x^{k+1} - x^*|}{|\delta_x^k|} = \alpha \cdot \frac{o(|\delta_x^k|)}{|\delta_x^k|},$$

which demonstrates Q-superlinear convergence.

For the converse, suppose s$\{x^k\} \to x^*$ Q-superlinearly. Using Lemma 2 and (3.5), we have that, for some $\eta > 0$,

$$|P_k(B_k - G_k)\delta_x^k + (P_{k+1} - P_k)(B_k - G_k)\delta_x^k| + |g(x^{k+1})| \leqq \eta |x^{k+1} - x^*|,$$

which, together with (3.4) and (3.6), imply that

$$\frac{|P_k(B_k - G_k)\delta_x^k|}{|\delta_x^k|} \leqq \eta \cdot \frac{|x^{k+1} - x^*|}{|\delta_x^k|} + O(|\delta_x^k|).$$

Again using Lemma 2, we have

$$\frac{|P_k(B_k - G_k)\delta_x^k|}{|\delta_x^k|} \leqq \eta \cdot \frac{|x^{k+1} - x^*|}{|\delta_x^k|} + O(|\delta_x^k|) \leqq \hat{\eta} \cdot \frac{|x^{k+1} - x^*|}{|x^k - x^*|} + O(|\delta_x^k|).$$

Letting $k \to \infty$ (and hence $|\delta_x^k| \to 0$) gives the desired results.

We note that if $f(x)$ is augmented by the penalty term $cg(x)^T g(x)$ with c a large positive constant, then the second order sufficiency conditions imply that the Hessian of the augmented Lagrangian is positive definite at (x^*, λ^*). Moreover, it is easily shown that the formula (1.9) is unchanged by this added term; thus the only effect is in the update formula (1.4). If, as is common, the B_k are chosen to approximate $l_{xx}(x^*, \lambda^*)$ in the sense that

$$B_{k+1}\delta_x^k = y^k \equiv l_x(x^{k+1}, \lambda^{k+1}) - l_x(x^k, \lambda^{k+1}), \tag{3.8}$$

then the assumption that $l_{xx}(x^*, \lambda^*)$ is positive definite makes the update formulas which preserve positive definiteness, such as the DFP or BFGS, natural candidates for use in this scheme. The next theorem shows that Q-superlinear convergence is achieved in these cases (cf. Han [6], Tapia [11], and Glad [5]). The following lemma is important for our proof.

LEMMA 4. *Let B_{k+1} be derived from B_k by either the* DFP *or the* BFGS *update with y^k given by* (3.8). *Assume that $l_{xx}(x^*, \lambda^*)$ is positive definite and $\{x^k\}$ converges linearly to x^*. Let B^*_{k+1} be generated from B_k using the same update formula as for B_{k+1} but with y^k replaced by y^*, where*

$$y^* \equiv l_x(x^{k+1}, \lambda^*) - l_x(x^k, \lambda^*).$$

Then

$$|B_{k+1} - B^*_{k+1}| \leqq \alpha\sigma((x^{k+1}, \lambda^{k+1}), (x^k, \lambda^k)), \tag{3.9}$$

where α is a constant independent of k and

$$\sigma((x^{k+1}, \lambda^{k+1}), (x^k, \lambda^k)) = \max\{|(x^j, \lambda^j) - (x^*, \lambda^*)|: j = k, k+1\}.$$

Proof. We prove the lemma for the BFGS update. The proof for the DFP update is similar but more laborious. From the definitions of y^k and y^* we have

$$y^k - y^* = (\nabla g(x^{k+1}) - \nabla g(x^k))(\lambda^{k+1} - \lambda^*),$$

and thus, there is a constant β_1 such that

$$|y^k - y^*| \leqq \beta_1 |\delta_x^k| |\lambda^{k+1} - \lambda^*|.$$

From the assumptions there exist positive constants η_1, η_2 such that for large k

$$(y^k)^T\delta_x^k \geqq \eta_1|\delta_x^k|^2, \qquad |y^k| \leqq \eta_2|\delta_x^k|,$$

$$(y^*)^T\delta_x^k \geqq \eta_1|\delta_x^k|^2, \qquad |y^*| \leqq \eta_2|\delta_x^k|.$$

For the BFGS update,

$$B_{k+1} - B^*_{k+1} = \frac{((y^*)^T\delta_x^k)y^k(y^k)^T - ((y^k)^T\delta_x^k)y^*(y^*)^T}{((y^*)^T\delta_x^k)((y^k)^T\delta_x^k)},$$

from which it follows that

$$|B_{k+1} - B^*_{k+1}| \leqq \frac{3|y^k||y^*||\delta_x^k||y^k - y^*|}{|(y^*)^T\delta_x^k||(y^k)^T\delta_x^k|} \leqq 3\beta_1(\eta_2/\eta_1)^2|\lambda^{k+1} - \lambda^*|.$$

Inequality (3.9) follows immediately.

THEOREM 3.2. *Assume $l_{xx}(x^*, \lambda^*)$ is positive definite. If the B_k are obtained by either the* DFP *or* BFGS *formulas with y^k defined by* (3.8) *and if $\{x^k\}$ converges to x^* linearly, then the convergence is Q-superlinear.*

Proof. The proof follows the lines of argument used in unconstrained optimization. Let y^* and B^*_{k+1} be as defined in Lemma 4. From our assumptions x^* is an unconstrained minimum of $l(x, \lambda^*)$ and hence the results of Broyden, Dennis and Moré [2] for the unconstrained case can be applied to obtain the fundamental inequality:

$$\begin{aligned}\|B^*_{k+1} - l_{xx}(x^*, \lambda^*)\|_M \leqq \{(1 - c\theta_k^2)^{1/2} + \alpha_1\hat{\sigma}(x^{k+1}, x^k)\}\|B_k - l_{xx}(x^*, \lambda^*)\|_M \\ + \alpha_2\hat{\sigma}(x^{k+1}, x^k),\end{aligned} \tag{3.10}$$

where α_1 and α_2 are constants independent of k,

$$\hat{\sigma}(x^{k+1}, x^k) = \max\{|x^{k+1} - x^*|, |x^k - x^*|\},$$

$$\theta_k = \frac{|M(B_k - l_{xx}(x^*, \lambda^*))\delta_x^k|}{\|B_k - l_{xx}(x^*, \lambda^*)\|_M |M^{-1}\delta_x^k|},$$

$$M = l_{xx}(x^*, \lambda^*)^{-1/2},$$

and the M-norm, $\|Q\|_M$, stands for the Frobenius norm of the matrix MQM. The triangle inequality can now be used with (3.9) and (3.10) to establish

$$\begin{aligned} \|B_{k+1} - l_{xx}(x^*, \lambda^*)\|_M \leqq \{(1 - c\theta_k^2)^{1/2} + \alpha_3\hat{\sigma}(x^{k+1}, x^k)\}\|B_k - l_{xx}(x^*, \lambda^*)\|_M \\ + \alpha_4\sigma((x^{k+1}, \lambda^{k+1}), (x^k, \lambda^k)), \end{aligned} \tag{3.11}$$

where B_{k+1} is the DFP and BFGS update of B_k and σ is defined as in Lemma 4. Since $\{x^k\}$ converges to x^* (and hence $\{\lambda^k\}$ converges to λ^*) it follows that $\{\|B_k - l_{xx}(x^*, \lambda^*)\|_M\}$ has a limit (Dennis and Moré [3]). If the limit is not zero, then (3.11) implies $\theta_k \to 0$; if the limit is zero then $\|B_k - l_{xx}(x^*, \lambda^*)\|_M \to 0$. In either case we have

$$\lim_{k\to\infty} \frac{|(B_k - l_{xx}(x^*, \lambda^*))\delta_x^k|}{|\delta_x^k|} = 0.$$

Since the projection matrices $P(x^k)$ are bounded, Theorem 3.1 can be applied to establish the Q-superlinear convergence.

Theorem 3.2 rests heavily on the assumption that $l_{xx}(x^*, \lambda^*)$ is positive definite. In theory, $l_{xx}(x^*, \lambda^*)$ need only be positive definite on the null space of $\nabla g(x^*)^T$. Nevertheless, most implementations of the quasi-Newton approach use updates (such as BFGS) which maintain positive definiteness of the B_k (with some ad hoc scheme to assure that $(y^k)^T\delta_x^k$ is positive). It remains an open question as to whether Q-superlinear convergence can be guaranteed with these approaches.

In the previous theorems, linear convergence of the $\{x^k\}$ is assumed. However, if the bounded deterioration inequality (3.11) holds, then linear convergence can be achieved by requiring $|x^0 - x^*|$ and $|B_0 - l_{xx}(x^*, \lambda^*)|$ to be sufficiently small. As shown above, (3.11) holds when $l_{xx}(x^*, \lambda^*)$ is positive definite. Without the positive definite assumptions the usual conditions for linear convergence require that $|B_k - l_{xx}(x^*, \lambda^*)|$ be small for all k. (See Han [6] and Tapia [10] for the relevant results.) In the next theorem we relax this restriction by showing linear convergence under the requirement that $|P(x^k)(B_k - l_{xx}(x^*, \lambda^{**}))|$ be small for all k. This theorem further illustrates the significance of the projection operator in the quasi-Newton theory for constrained minimization.

THEOREM 3.3. *Let the B_k satisfy*

$$|B_k^{-1}| \leqq \eta$$

for some $\eta > 0$. Then there exist positive constants ε and ξ such that if

(i) $$|x^0 - x^*| < \xi,$$

(ii) $$|P(x^*)(B_k - l_{xx}(x^*, \lambda^*))| < \varepsilon \quad \textit{for all } k \geqq 0,$$

then the sequence $\{x^k\}$ generated by

$$x^{k+1} = x^k - B_k^{-1} l_x(x^k, \Lambda_k(x^k)), \tag{3.12}$$

where

$$\Lambda_k(x) = (\nabla g(x)^T B_k^{-1} \nabla g(x))^{-1}(g(x) - \nabla g(x)^T B_k^{-1} \nabla f(x)), \tag{3.13}$$

is well defined and converges linearly to x^*.

Remark. The iteration (3.12)–(3.13) is equivalent to (1.6)–(1.7), but this form makes the proof easier.

Proof. As demonstrated earlier, it follows from the assumptions that for some $\hat{\xi} > 0$ and $|x - x^*| < \hat{\xi}$, $(\nabla g(x)^T B_k^{-1} \nabla g(x))^{-1}$ exists and is uniformly bounded. Thus, for $|x^0 - x^*| < \hat{\xi}$, x^1 is well defined. Since $\Lambda_k(x^*) = \lambda^*$ for all k, we have

$$\begin{aligned} x^1 - x^* &= x^0 - x^* - B_0^{-1} l_x(x^0, \Lambda_0(x^0)) \\ &= B_0^{-1}\{B_0 - l_{xx}(x^*, \lambda^*) - \nabla g(x^*)\nabla \Lambda_0(x^*)^T\}(x^0 - x^*) + h^0(x^0), \end{aligned}$$

where $\nabla \Lambda_0(x^*)$ denotes the Jacobian of Λ_0 at $x = x^*$ and $|h^0(x^0)| \leqq \alpha^0 |x^0 - x^*|^2$, α^0 constant. From (3.13) we see that

$$\nabla \Lambda_0(x^*)^T = (\nabla g(x^*)^T B_0^{-1} \nabla g(x^*))^{-1} \nabla g(x^*)^T B_0^{-1}(B_0 - l_{xx}(x^*, \lambda^*)).$$

Therefore,

$$\begin{aligned} |x^1 - x^*| \leqq |B_0^{-1}| \cdot |\{I - \nabla g(x^*)(\nabla g(x^*)^T B_0^{-1} \nabla g(x^*))^{-1} \nabla g(x^*)^T B_0^{-1}\} \\ \cdot (B_0 - l_{xx}(x^*, \lambda^*))| \cdot |x^0 - x^*| + \alpha^0 |x^0 - x^*|^2. \end{aligned}$$

Let $V_k^* = I - \nabla g(x^*)(\nabla g(x^*)^T B_k^{-1} \nabla g(x^*))^{-1} \nabla g(x^*)^T B_k^{-1}$ and note that as in (3.2b), $V_k^* P(x^*) = V_k^*$. Thus

$$|x^1 - x^*| \leqq |B_0^{-1}| \cdot |V_0^*| \cdot |P(x^*)(B_0 - l_{xx}(x^*, \lambda^*))| \cdot |x^0 - x^*| + \alpha^0 |x^0 - x^*|^2.$$

From our assumptions, it now follows that the $|V_k^*|$ will be uniformly bounded by, say, $\hat{\beta} > 0$. We now choose ε and ξ small enough so that $\eta\hat{\beta}\varepsilon + \alpha^0 \xi \leqq \rho < 1$, and therefore, $|x^1 - x^*| \leqq \rho |x^0 - x^*|$. The desired result can now be proven by induction since the sequence $\{\alpha^k\}$ can be uniformly bounded.

We observe that in the above theorem, condition (ii) could be replaced by

$$|P(x^k)(B_k - l_{xx}(x^*, \lambda^*))| < \varepsilon,$$

which is consistent with the form in Theorem 3.1.

Acknowledgment. The authors would like to thank R. A. Tapia for his helpful comments.

REFERENCES

[1] P. T. BOGGS AND J. W. TOLLE, *Augmented Lagrangians which are quadratic in the multiplier*, J. Opt. Theory Appl., 31 (1980), pp. 17–26.

[2] G. BROYDEN, J. DENNIS AND J. MORÉ, *On the local and superlinear convergence of quasi-Newton methods*, J. Inst. Math. Appl., 12 (1973), pp. 223–246.

[3] J. DENNIS AND J. MORÉ, *A characterization of superlinear convergence and its application to quasi-Newton Methods*, Math. Comp., 28 (1974), pp. 549–560.

[4] ———, *Quasi-Newton methods, motivation and theory*, SIAM Review, 19 (1977), pp. 46–89.

[5] S. GLAD, *Properties of updating methods for the multipliers in augmented Lagrangians*, J. Opt. Theory Appl., 28 (1979), pp. 135–156.

[6] S. P. HAN, *Dual variable metric algorithms for constrained optimization*, this Journal, 15 (1977), pp. 546–565.

[7] J. ORTEGA AND W. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[8] M. J. D. POWELL, *A fast algorithm for nonlinearly constrained optimization calculations*, 1977 Dundee Conference on Numerical Analysis, June, 1977.

[9] ———, *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, Nonlinear Programming 3, O. Mangasarian, R. Meyer and S. Robinson, eds., Academic Press, New York, 1978, pp. 27–63.

[10] R. A. TAPIA, *Diagonalized multiplier methods and quasi-Newton methods for constrained optimization*, J. Opt. Theory Appl., 22 (1977), pp. 135–194.

[11] ———, *Quasi-Newton methods for equality constrained optimization: Equivalence of existing methods and a new implementation*, Nonlinear Programming 3, O. Mangasarian, R. Meyer and S. Robinson, eds., Academic Press, New York, 1978, pp. 125–164.